Multitask Learning and its Applications to Alzheimer's Disease Progression and Cancer Survival Prediction

André Ricardo Gonçalves
Ph.D.

May 23, 2018

LLNL-PRES-751581
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

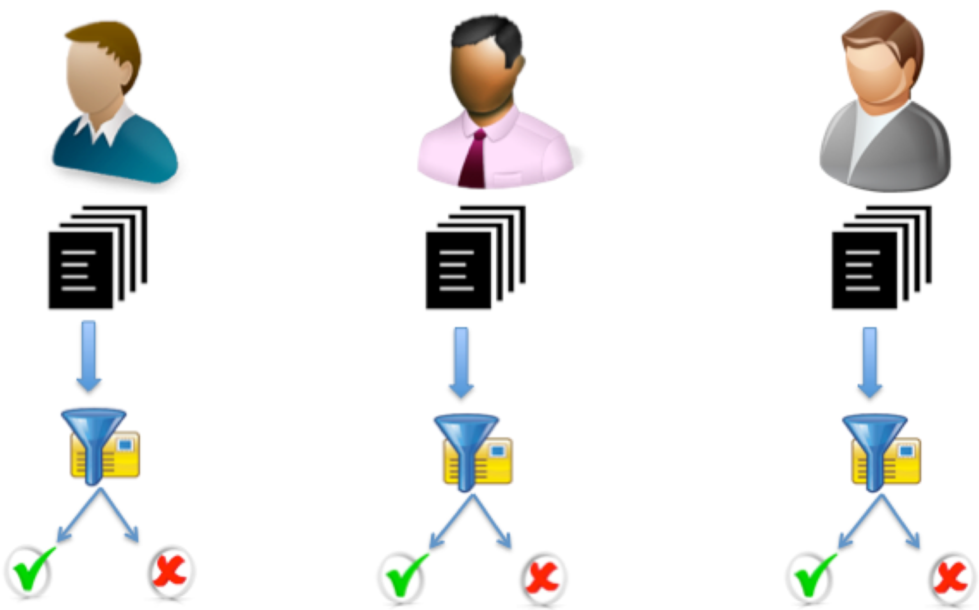
Lawrence Livermore National Laboratory

Outline

- Motivation
- Multitask Learning
  - Comparison with related areas
- Applications
  - Alzheimer's Disease progression
  - Cancer survival prediction
- Concluding remarks

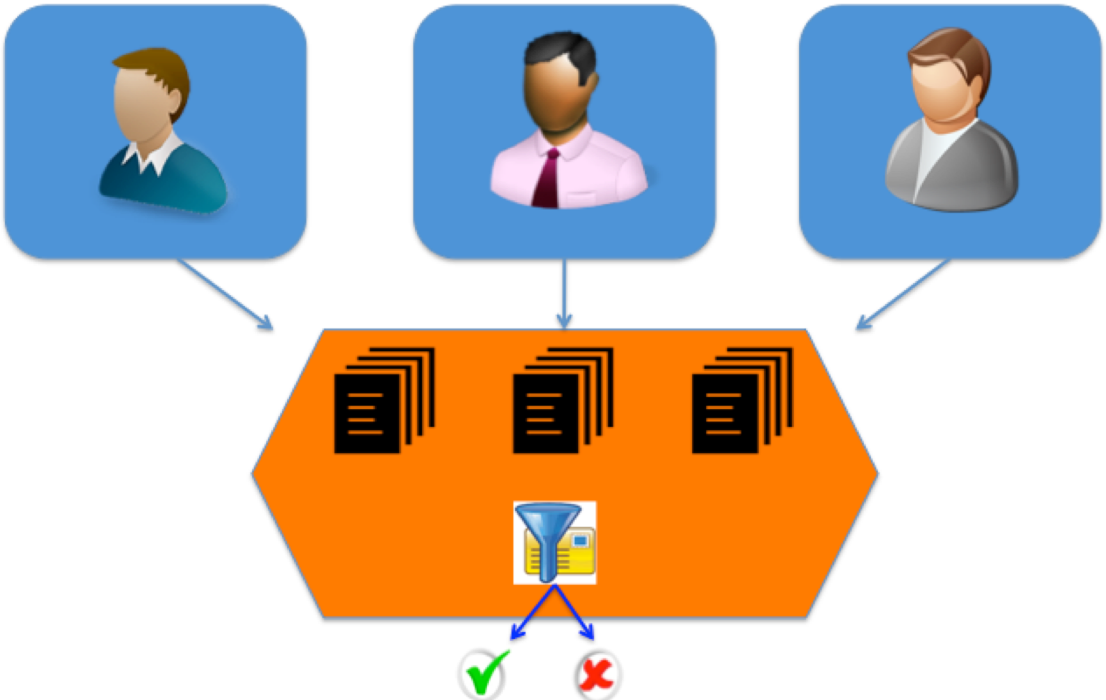
Learning multiple tasks

- Problem: Spam detection



Learning multiple tasks

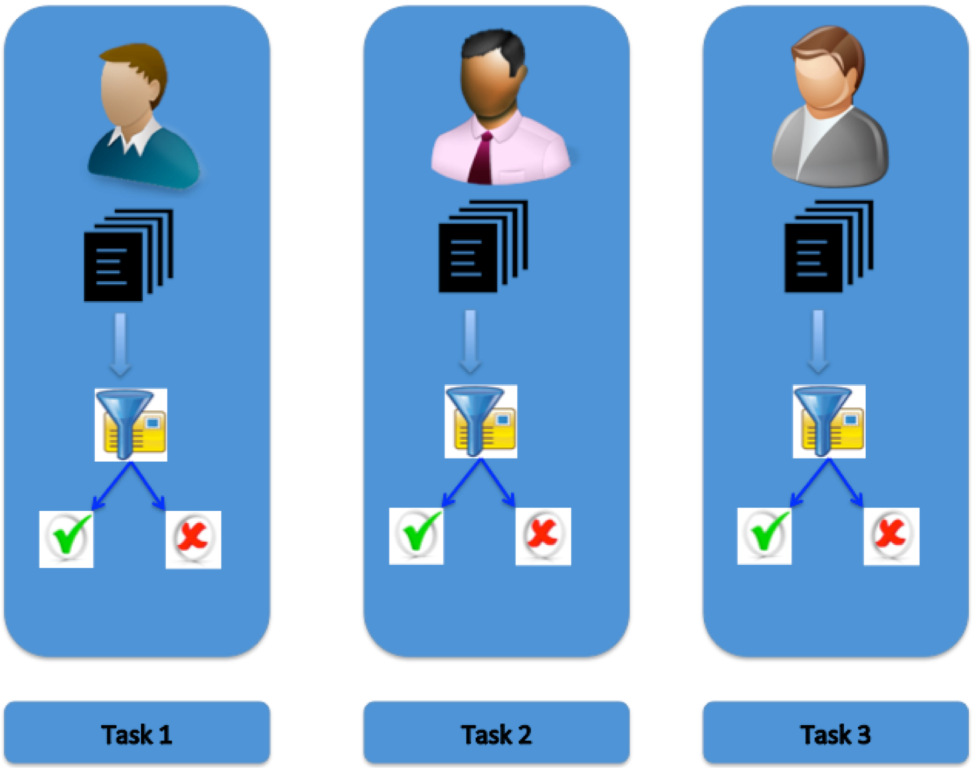
- Approach 1: single filter (classifier) for all users



- Deficiency: some users may have different behaviors ☹

Learning multiple tasks

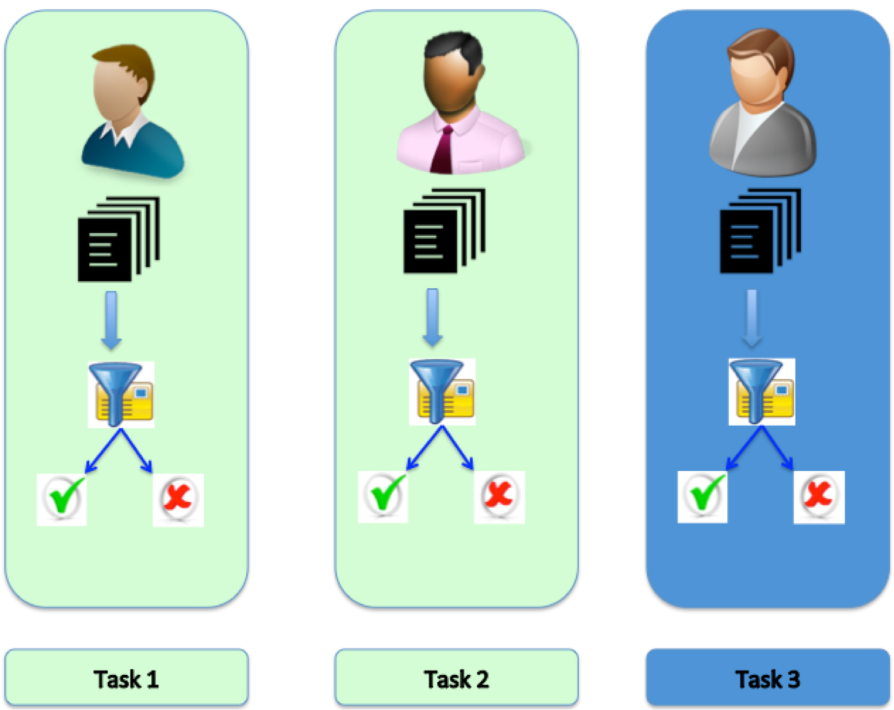
- Approach 2: One filter (classifier) for each user

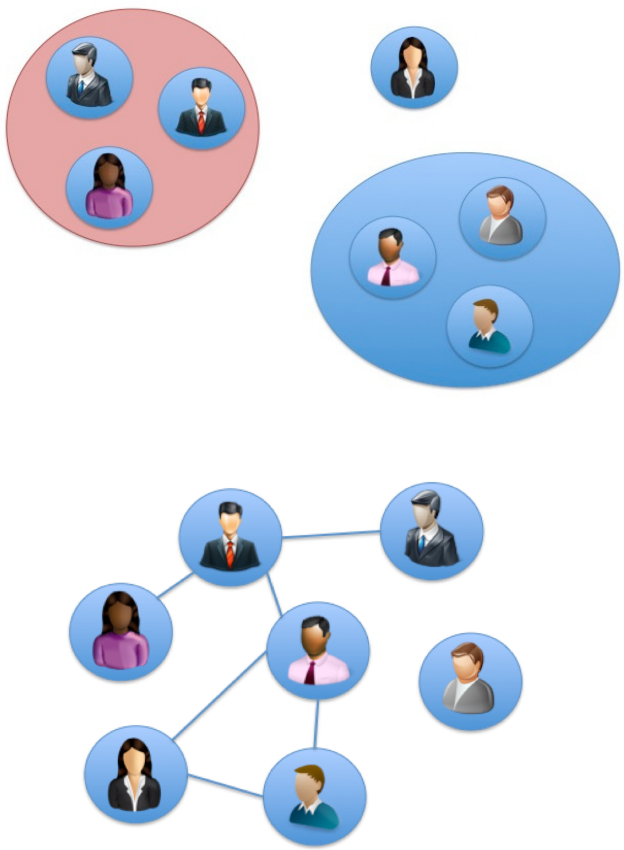


- Deficiency: some users may have similar behaviors and we are ignoring it ☹

Learning multiple tasks

- Multitask Learning: learn all tasks simultaneously while taking tasks relationship into the learning process





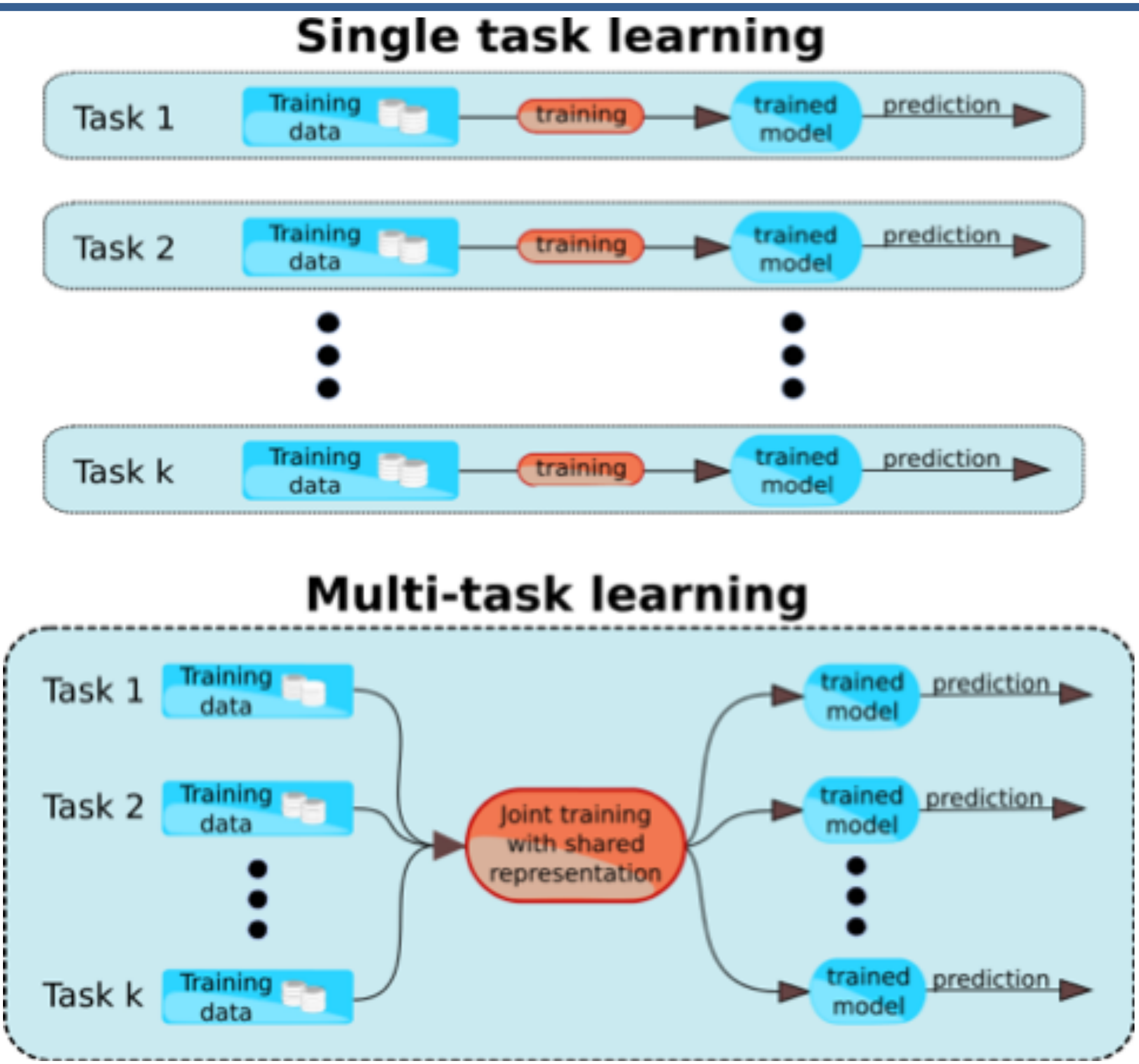
Learning multiple tasks

Definition:

Multitask Learning (MTL) is a machine learning paradigm which seeks to improve the generalization of a learning task by using auxiliary information from another related tasks [Caruana, 1997].

Unlike traditional learning, where each task is learned independently, here all tasks are learned simultaneously and information about its relationship is also considered.

Traditional single task learning vs MTL



Mathematical definition of MTL

Summation over all tasks

$$\min_{\Theta} \sum_{k=1}^{m} \left(\sum_{i=1}^{n_k} \ell(f_k(\boldsymbol{\theta}_k, \mathbf{x}_k^i), y_k^i) \right) + \mathcal{R}(\Theta)$$

Regularization to enforce tasks sharing

Empirical risk for the k-th task

Other areas related to MTL

- Multiple-Output Regression:

$$J(\Theta) = \sum_{k=1}^{m} \frac{1}{n} \sum_{i=1}^{n} (y_k^i - \boldsymbol{\theta}_k^\top \mathbf{x}_i)^2.$$

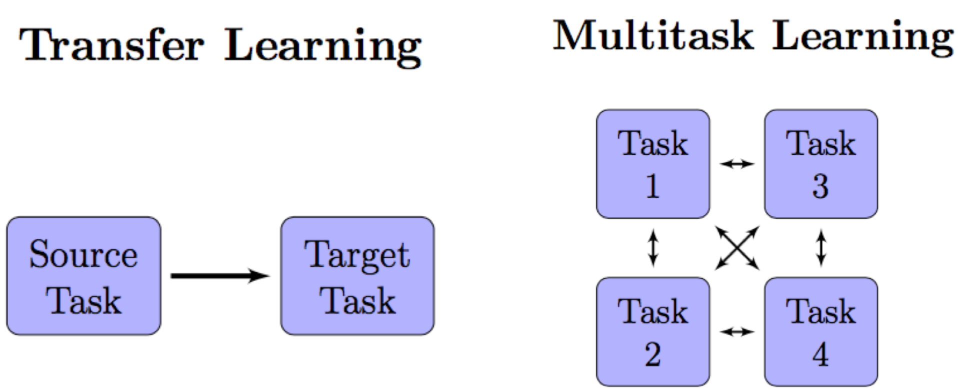
Differences:

a) **Independent single response regression model (no regularization added);**

b) **Inputs (covariates) are the same for all regressors (X_1 = X_2 = … = X_m)**

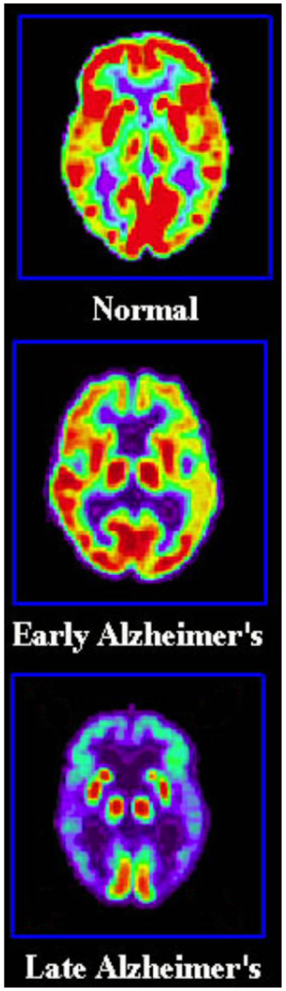
- Multilabel Classification:
  - Similar to multiple-output regression, but for classification problems.
  - Binary relevance transformation

- Transfer Learning:



Assessing Alzheimer’s Disease Progression

- Alzheimer’s Disease (AD) is a severe neurodegenerative disorder that results in loss of mental function due to the deterioration of brain tissue [Khachaturian, 1985]
- Early diagnosis of AD is key to the development, assessment, and monitoring of new treatments for AD
- Cognitive scores are used to measure patient cognitive capabilities such as attention, memory, language and visuo-constructional functions
- Disease stage can be characterized based on such cognitive score

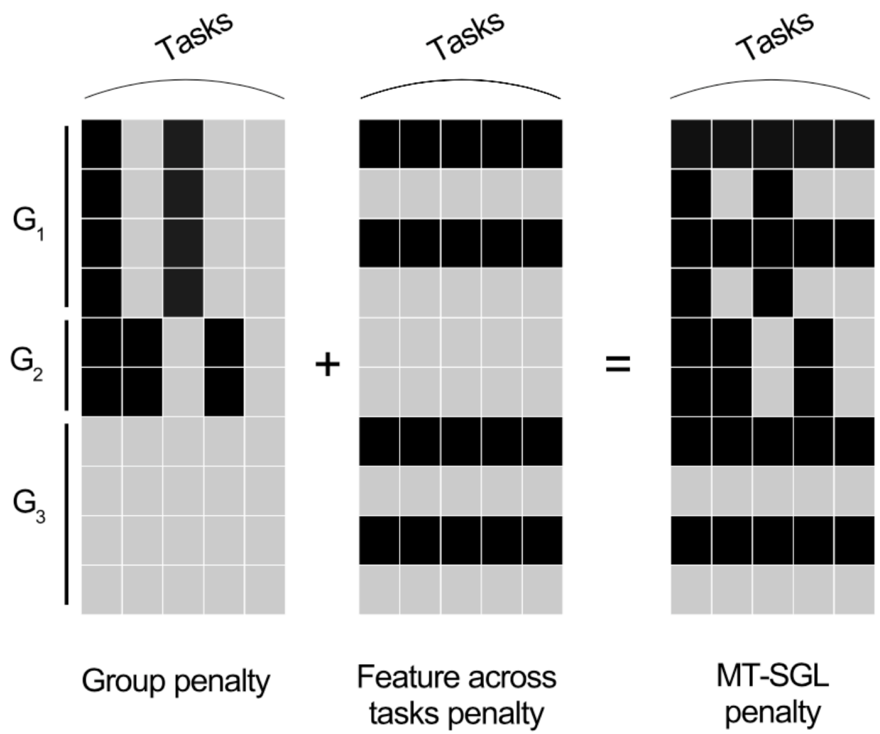


Assessing Alzheimer's Disease Progression with MTL

- Jointly prediction of five cognitive scores based on properties of regions-of-interest (ROIs) in the brain
- Considered cognitive scores:
  - Alzheimer's Disease Assessment Scale – cognitive total score (ADAS)
  - Mini Mental State Exam score (MMSE)
  - Rey Auditory Verbal Learning Test (RAVLT)
    - Total score (RAVLT-TOTAL)
    - 30 minutes delay score (RAVLT-30)
  - RAVLT recognition score (RECOG)
- Magnetic resonance imaging (MRI), positron emission tomography (PET), along with other biomarkers
- Data from "The Alzheimer's Disease Neuroimaging Initiative" (ADNI)

Assessing AD Progression: MTL model

- MTL with sparse group-structured penalty (MT-SGL)

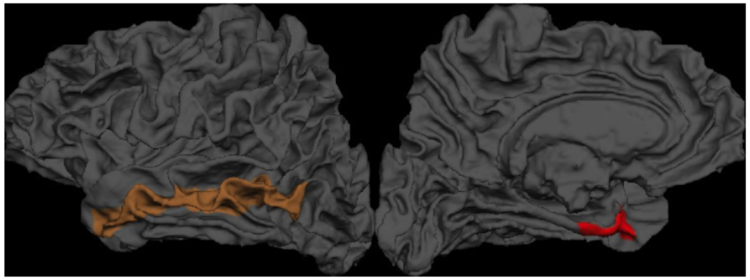


X. Liu, A.R. Goncalves, P. Cao, D. Zhao, A. Banerjee. Modeling Alzheimer's disease cognitive scores using multi-task sparse group lasso. *Computerized Medical Imaging and Graphics,* vol. 66, 2018.

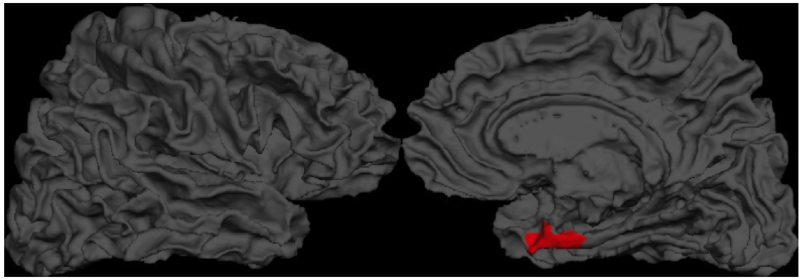
AD Progression with MTL: Results

- 816 patients, 92 ROIs (1 to 4 features per ROI): 327 features
- Results (relative RMSE improvement):
  - Over Ridge Regression: ~20%
  - Over Group Lasso: 5% ~ 6%
  - Improvement over other MTL methods
- Researchers are not only interested in more accurate cognitive scores prediction:
  - Identify the brain areas more affected by the disease at each stage
- Interpretable models are preferred

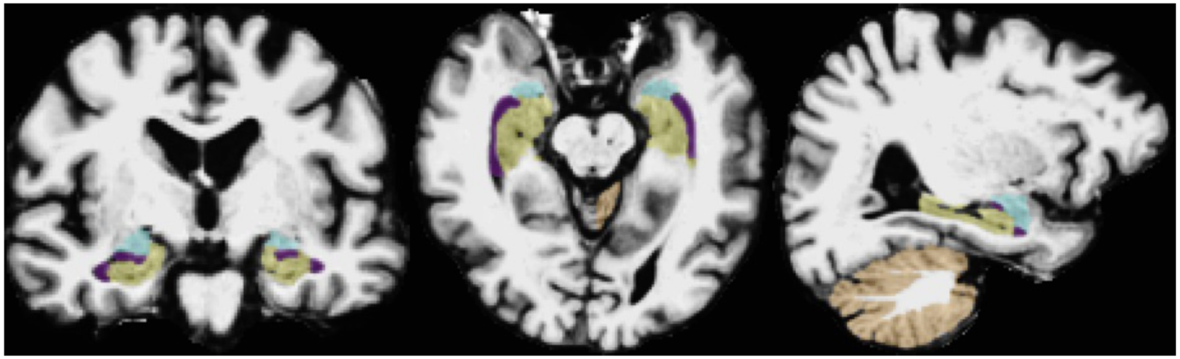
Relevant ROIs identified by MT-SGL



(a) Left-Hemisphere: *Left Middle Temporal* and *Left Entorhinal*.



(b) Right-Hemisphere: No relevant ROI selected and *Right Entorhinal*.



(c) Subcortical: *Left Inferior Lateral Ventricle, Left Cerebellum Cortex, Left Hippocampus, Left Amygdala, Right Inferior Lateral Ventricle, Right Hippocampus, Right Amygdala*.

Cancer Survival Prediction

- Goal: Leverage MTL to share information across anatomically distinct cancer types
- Our hypothesis is that as stark as the differences between cancers may be, there exist underlying processes that are common, even across disparate cancer types
- We investigate whether MTL classifiers can provide improved predictions of 5-year survival

A.R. Goncalves, A.P. Sales, B. Soper, M. Nygard, J. Nygard. Improving Survival Predictions via Multitask Learning Across Cancer Types. Machine Learning for Health Care. 2018 (Submitted)

Cancer Survival Prediction: Data

- SEER public dataset (from National Cancer Institute – NCI)
- Only cancers topologies for which human papillomavirus infection (HPV) is a known risk factor
- Considered cases diagnosed during the period of 2004 to 2014, totaling 27586 cancer cases
- Dataset consists of 16 covariates (mostly categorical) and the binary outcome tells if a patient has survived for at least 5 years
- Data is right-censored: patients dropped out the study before 5-years of follow-up
- Tasks were split in three different manners, based on anatomical groupings of cancer types.

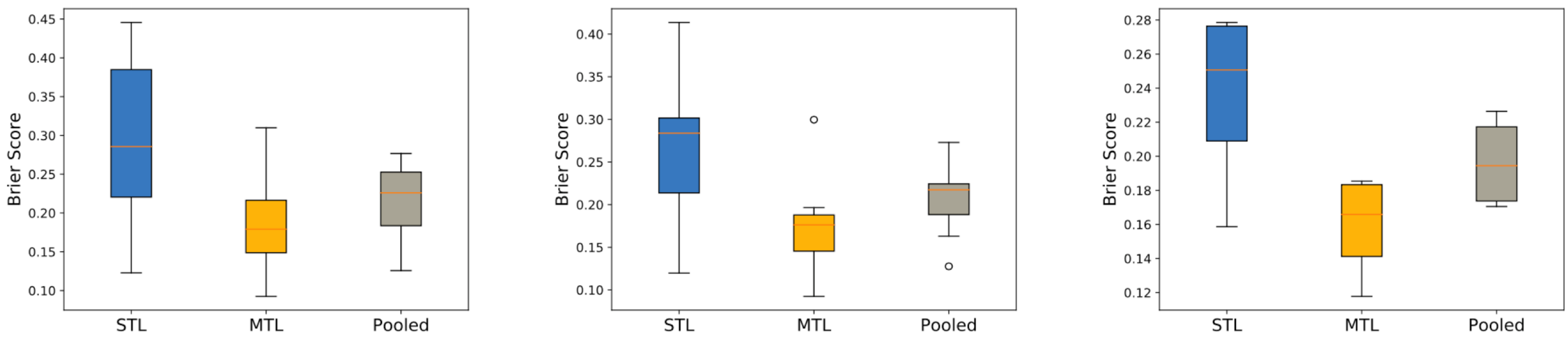
Cancer Survival Prediction: MTL model

- Model is an extension of the model proposed in *Goncalves et. al* 2016, to handle right-censored data using IPCW adaptation

$$\mathbf{W} = \underset{\mathbf{W},\mathbf{\Omega}\succeq 0}{\arg\min} \sum_{t=1}^{T} \frac{1}{n_t} \sum_{i=1}^{n_t} \omega_i^t \mathcal{L}(y_i^t, \mathbf{x}_i^t, \mathbf{w}_t) + \lambda_1 \mathrm{tr}(\mathbf{W}\mathbf{\Omega}\mathbf{W}^T) - d \log |\mathbf{\Omega}| + \lambda_2 \|\mathbf{\Omega}\|_1$$

- Jointly learn the matrix of tasks parameters (W) and a matrix that captures the dependence among tasks (Omega)
- Weights ω_i^t are computed by IPCW to account for censored data
- MTL model is compared against STL and Pooled models

Cancer Survival Prediction: Results



(a) (b) (c)

Figure 1: 'Per-task basis' comparison. Brier scores for STL, MTL, and pooled classifiers for the three task definitions: (a) Level 1 (11 tasks), Level 2 (7 tasks), Level 3 (4 tasks). Each boxplot contains as many points as there are tasks, with each point being the mean Brier score for that task over the 30 train/test runs.

Concluding remarks

- Multitask learning is a powerful machine learning paradigm
- Many machine learning problems can be cast to a MTL problem
- MTL benefits the most when multiple related models are to be trained and relative small (considering model complexity) datasets are available

Lawrence Livermore
National Laboratory